DOCUMENT RESUME

ED 413 353 TM 027 682

| | |
|---|---|
| AUTHOR | Bell, Gregory |
| TITLE | Making Appropriate & Ethical Choices in Large-Scale Assessments: A Model Policy Code. |
| INSTITUTION | North Central Regional Educational Lab., Oak Brook, IL. |
| SPONS AGENCY | Office of Educational Research and Improvement (ED), Washington, DC. |
| REPORT NO | RPIC-MAEC-94 |
| PUB DATE | 1994-12-04 |
| NOTE | 33p. |
| CONTRACT | RP91002007 |
| AVAILABLE FROM | North Central Regional Educational Laboratory, 1900 Spring Road, Suite 300, Oak Brook, IL 60521 ($15.95). |
| PUB TYPE | Guides - Non-Classroom (055) |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | Codes of Ethics; *Educational Assessment; Educational Policy; Educational Testing; Elementary Secondary Education; *Ethics; Evaluation Methods; Models; *Policy Formation; *Scoring; Test Bias; Test Coaching; Test Construction; Test Interpretation; *Test Use |
| IDENTIFIERS | *Large Scale Assessment; Large Scale Programs |

ABSTRACT

This set of policy statements is intended to provide guidance to those who evaluate and select assessments, prepare students for those assessments, administer and score the tests, and interpret and use assessment results to make decisions about students and schools. The focus is on large-scale assessments that have consequences for students and schools. The fundamental principles of appropriate and ethical assessment practice are reviewed. They center around the fundamental reason for assessment, promoting the education of students through accurate measurement of their learning. Guidelines for evaluating and selecting assessments take into account the importance of choosing an assessment that is appropriate for its intended purposes and then ensuring that its use will be fair. The preparation of students for an assessment is the source of many testing problems. Guidelines for this area center on appropriate test preparation that does not give unfair advantages or compromise the validity of the test results. Recommendations for administering and scoring assessments refer to testing conditions and monitoring practices on test day, as well as fairness in scoring. Recommendations are also made for the interpretation and use of test results in an ethical manner. Appendix A defines terms used in the Code and discussion, and Appendix B is a bibliography. (Contains 25 references.) (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 413 353

# *Making Appropriate & Ethical Choices in Large-Scale Assessments*

## *A Model Policy Code*

NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY

TM027682

*By Gregory Bell*

*December 4, 1994*

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

NCREL



BEST COPY AVAILABLE

**North Central Regional Educational Laboratory**
1900 Spring Road, Suite 300
Oak Brook, IL 60521
(708) 571-4700, Fax (708) 571-4716

Jeri Nowakowski: *Executive Director*

Deanna H. Durrett: *Regional Policy Information Center (RPIC) Director*

Lawrence B. Friedman: *Associate Director, RPIC*

Linda Ann Bond: *Director of Assessment, RPIC*

NCREL is one of ten federally supported educational laboratories in the country. It works with education professionals in a seven-state region to support restructuring to promote learning for all students—especially students most at risk of academic failure in rural and urban schools.

The Regional Policy Information Center (RPIC) connects research and policy by providing federal, state, and local policymakers with research-based information on such topics as educational governance and student assessment policy.

© 1994 North Central Regional Educational Laboratory

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002007. The content of this publication does not necessarily reflect the views of OERI, the Department of Education, or any other agency of the U.S. Government.

RPIC-MAEC-94 $15.95

# Making Appropriate & Ethical Choices in Large-Scale Assessments

## A Model Policy Code

by Gregory Bell

## TABLE OF CONTENTS


Introduction . . . . . 1

A Model Policy Code . . . . . 3

A. Fundamental Principles of Appropriate and Ethical Assessment Practice . . . . . 3

B. On Evaluating and Selecting Assessments . . . . . 4
1. Select an assessment instrument that is appropriate for its intended purpose(s) and populations(s) . . . . . 4
2. Select assessments that minimize opportunities for misuse or misinterpretation of results . . . . . 6
3. Insist that assessment developers comply with their responsibilities to those evaluating their instruments for selection . . . . . 6
4. Recognize and avoid the mixed motives and conflicts of interest that may be present in the development and selection of an assessment . . . . . 7
5. Establish procedures to ensure that assessments being developed or evaluated are kept confidential in order to minimize disclosure of actual assessment content, questions or their format, or other information which could provide advantage in student preparation . . . . . 7

C. On Preparing Students for an Assessment . . . . . 8
1. Follow general guidelines for preparing students for an assessment . . . . . 9
2. Avoid test-specific instruction when preparing students for an assessment . . . . . 9
3. Provide for monitoring and identification of inappropriate preparation activities . . . . . 10
4. Follow guidelines on other specific preparation activities . . . . . 10

D. On Administering and Scoring Assessments . . . . . 11
1. Be aware of general guidelines on assessment administration . . . . . 11
2. Provide testing conditions that will not undermine the reliability and validity of an assessment . . . . . 13
3. Limit access to assessment materials before, during and after its administration . . . . 14
4. Monitor assessment administration . . . . . 15
5. Monitor assessment scoring to ensure the reliability of its results . . . . . 15
6. Be aware of guidelines on other specific, administration activities . . . . . 16

E. On Interpretation and Use of Assessment Results . . . . . 16
1. Refrain from inappropriate and improper interpretations and uses of assessment results . . . . . 17
2. Promote appropriate interpretation and use of an assessments' results by encouraging valid inferences . . . . . 17
3. Provide information to students and their representatives about access to the assessment results and steps taken to protect their privacy . . . . . 18

4. Ensure that test developers fulfill their obligation to assist in proper communication of assessment results . . . . . 18
5. Identify all audiences to which the results of an assessment may be communicated and educate them in appropriate interpretations and uses of its results . . . . . 19
6. Minimize potential misinterpretation and misuse when reporting assessment results to students, parents, legal representatives, teachers, the general public, and the media . . . . . 19
7. Communicate the results of an assessment, and interpretations of the results, within the context of the assessment's limitations . . . . . 19

Conclusion . . . . . 20

APPENDIX A . . . . . 21

APPENDIX B . . . . . 23

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521
(708) 571-4700, Fax (708) 571-4716

Jeri Nowakowski: Executive Director

Deanna H. Durrett: Director, RPIC

Lawrence B. Friedman: Associate Director, RPIC

Linda Ann Bond: Director of Assessment, RPIC

John Blaser: Editor

NCREL is one of ten federally supported educational laboratories in the country. It works with education professionals in a seven-state region to support restructuring to promote learning for all students–especially students most at risk of academic failure in rural and urban schools.

The Regional Policy Information Center (RPIC) connects research and policy by providing federal, state, and local policymakers with research-based information on such topics as educational governance, teacher education, and student assessment policy.

1994 North Central Regional Educational Laboratory

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002007. The contents of this publication do not necessarily reflect the views of OERI, the Department of Education, or any other agency of the U.S. Government.

RPIC-MA-94

*Introduction*

Student assessment[1] continues to take center stage in the educational policy debate. Public cries for assessment systems that would hold all students, schools, and local districts accountable for student achievement echo throughout the nation. Indeed, assessment plays a pivotal role in 45 state educational reform strategies. However, when the results of external student assessments are used to judge the quality of students and schools,[2] the pressure to do well on the test can lead to inappropriate and unethical choices when tests are selected, prepared for, and administered, and when their results are interpreted and used. Simply stated, any assessment practice that results in differences in assessment results that are not due to differences in student knowledge and skills affects the accuracy of the assessment itself, and thus undermines the decisions based upon those results.

The problems that arise through inappropriate large-scale assessment practice can result in dire consequences for students and schools. If school A, for example, provides coaching to students before and during an assessment, giving students clues to the "right" answer, while school B does not, school B's students may be inappropriately identified as below-standard when their performance is compared to the inappropriately inflated performance of students in school A. When important decisions, such as student advancement or teacher evaluation, are dependent upon the assessment's results and comparisons of the results among students and schools are being made, the fundamental unfairness of this situation is clear.

If the integrity of an assessment is undermined, the distinction between the terms "inappropriate" and "unethical," as used in this Code, for the most part fades. This conclusion is clear once one considers the goal of this document—outlining the responsibilities of those in assessment and describing assessment practices that are both appropriate and inappropriate, so that the reliability and validity of an assessment program is not undermined. If the integrity of an assessment is compromised, the inferences drawn from its results may be corrupted. Decisions taken from these inferences may have a wide-ranging, and sometimes devastating, impact on individuals and institutions. When based on inferences derived from incomplete or erroneous data,

---

[1] Student assessment refers to any method of determining what students know and can do, including testing, teacher observations, collections of students' work (portfolios), projects, and performances. Distinctions are made between traditional assessments, which require students to select a right answer, and performance assessments, which require students to create an appropriate response to a performance task. The most common traditional assessments include multiple-choice, true-false, and matching tests. Performance assessments can range from essay exams to situations where students must build a model or design and perform an experiment. In the first case, student performance is inferred from the test score, while in the second case, the performance is directly observed.

[2] This policy deals specifically with the ethical issues surrounding large-scale, high stakes assessments created by districts, states, or nations for evaluative purposes.

these decisions, and their impacts, may be without foundation. If the decisions adversely affect individuals or institutions, the effect is thus inherently unfair. Moreover, inaccurate measurement of achievement, especially when it portrays low achievement, can have harmful and long-lasting effects on an individual student's life. The inescapable conclusion from all of this is that engaging in inappropriate assessment practices undermines the reliability (i.e., consistency) and validity (i.e., accuracy as an indicator of student performance on the content assessed) of the testing program and should be considered unethical.

While there are a great many areas where wrong choices can be made, this set of policy statements is intended to provide guidance to those who evaluate and select assessments, prepare students for those assessments, administer and score the assessments, and interpret and use assessment results to make decisions about students and schools. We focus on large-scale assessments; that is, district, state and national assessments that have consequences for students and schools. Although accurate results are also important in the classroom, a teacher may exert his or her professional judgment to determine whether or not a given classroom assessment result can be relied upon, and can amend the decisions that are made appropriately. Large-scale assessments do not provide this flexibility.

It is likely that inappropriate choices in assessment are often made unintentionally, resulting from actions by individuals who are either unaware of, or do not completely understand, their roles or the range of appropriate practices and uses of standardized testing. A better understanding may provide those involved with assessments with a foundation upon which to make informed decisions. To that end, this policy Code attempts to identify the chief responsibilities of educators, administrators, and education officials at the classroom, district and state levels in conducting an assessment. These responsibilities are divided in the Code by function including evaluating and selecting an assessment instrument, preparing students for testing, administering the assessment itself and preparing its surrounding environment, and interpreting or using results of the testing program. They are derived from various codes and guidelines that have been advanced by the professional assessment community and from the literature addressing these issues.[3]

It should be emphasized that questions surrounding appropriate and inappropriate testing practices should remain open to discussion among <u>all</u> who are involved in assessment. The potential for inappropriate or unethical testing practices will be greatly reduced if those developing assessment policy involve all who have a role in testing to assist them in their undertaking at each step, and engage them in a dialogue on the issues surrounding appropriate and ethical assessment practices. Such a continuing conversation will at least help to foster greater understanding at all levels—from assessment professionals to policymakers to classroom teachers—of where testing practices cross the line. We hope that this Code will assist in encouraging these conversations.

---

[3] The sources consulted in compiling this Code are listed below in Appendix B. This model policy code is based upon a previous paper by the author, *The Test of Testing: Making Appropriate and Ethical Choices in Assessment* (1993). The paper provides more extensive discussion of the issue of appropriate and ethical practice in large-scale assessments.

## The Key Terms

Although definitions of terms used in this Code may be found in Appendix A, several key terms need to be clarified now. *Assessment*, "the process of collecting, synthesizing and interpreting information for use in decision-making" (Airasian, 1991), is an essential tool for anyone making educational decisions about students or schools. The quality of these decisions will only be as good as the assessment data, and the interpretation of that data, used during the decision-making process. *Large-scale assessment* refers to assessments that are used in a number of different settings or schools, typically mandated at the district, state, or national level. This can be contrasted with classroom assessment, which takes place in a classroom and is most often administered by the teacher. This Code applies most particularly to large-scale assessments. Either category may employ different types of assessment strategies (for example, traditional paper-and-pencil tests requiring students to select a right answer, or performance assessments requiring students to produce an appropriate response). Multiple-choice, true-false, and matching tests are among the most typical traditional assessments, while extended essay exams and presentations or projects are typical performance-based assessments. Appropriate and ethical practice is equally important in all types of assessment; the damage caused by abuse is quite severe.

## *A Model Policy Code*

### A. Fundamental Principles of Appropriate and Ethical Assessment Practice

The principles set forth below are not meant to be exhaustive, but rather to suggest a foundation for building a climate where appropriate and ethical assessment practices may flourish. The principles are derived for the most part from a draft of the *Code of Ethical Assessment Practices in Education*, developed by the National Council on Measurement in Education's Ad Hoc Committee on the Development of a Code of Ethics. All involved in large-scale assessment should:

Maintain a focus on the fundamental reason for an assessment—promoting the education of students through accurate measurement of their learning.

Ensure that choices made in assessment practice are consistent with one's obligation to act with honesty, integrity, due care, and fundamental fairness to all involved in the assessment or affected by its results.

Ensure that the assessments used are reliable and that no practice used in selecting, preparing for, or administering an assessment detracts from its validity.

Ensure that the decision made about students are based on assessment results that are an accurate indicator of the student's level of achievement.

Acquire the knowledge and skills necessary to their roles in assessments.

Help others (including other educators, administrators, parents, the media, and the general public) to understand appropriate assessment practices.

Promote appropriate and ethical assessment practices and encourage "ownership" of an assessment by placing value on its producing accurate and reliable results.

B. **On Evaluating and Selecting Assessments**

The development of large-scale assessments, and the process of selecting them, is an important area where choices made by developers, policy-makers and assessment administrators can have a significant impact on the usefulness of the assessment, and the validity of its results. Those involved in evaluating and selecting assessments have important responsibilities to see that they select instruments that are well crafted and suited for those who are to be assessed. They also must ensure that assessment developers live up to their obligation to provide the means and information necessary for the evaluation to be done accurately and effectively.

The provisions of this section were derived in large part from the *Code of Fair Testing Practices* (Joint Committee on Testing Practices), a draft of the *Code of Ethical Assessment in Education* (National Council on Measurement in Education [NCME] Ad Hoc Committee on the Development of a Code of Ethics), and an NCME task force report entitled *Regaining Trust: Enhancing the Credibility of School Testing Programs*.

When engaged in the process of evaluating and selecting an assessment, it is important to:

1. **Select an assessment instrument that is appropriate for its intended purpose(s) and population(s).**

    a. Define why an assessment is needed, clearly articulating its purpose(s) and expected use(s).

    b. Identify the characteristics of the population(s) that the assessment is meant to measure (e.g., cultural, socioeconomic, and other demographic factors).

    c. Evaluate each potential assessment instrument within the context of its intended purpose and expected use. Evaluators should:

        (1) Base evaluations of potential assessment instruments on existing evidence of their technical quality and utility.

(2) Corroborate developers' claims concerning their assessment instrument by consulting other useful sources, such as:

- Buros Mental Measurements Yearbooks
- Specimen sets, disclosed tests and sample questions, directions, manuals, answer sheets, and score reports
- Independent evaluations of the assessment
- Those who have used the assessment previously

(3) Evaluate evidence of a potential instrument's validity and reliability, the age and adequacy of norms used in its development, and whether any bias can be detected.

(4) Choose assessment instruments that minimize possible bias based on gender, ethnicity, socioeconomic status, religion, age, disability, and other relevant characteristics of the population(s) tested by:

(a) Seeking evidence from assessment developers and others in order to substantiate claims that the instrument minimizes potential bias.

(b) Evaluating evidence that the assessment may be validly administered, interpreted, and used for the population(s) to be assessed. Evaluation should include both the assessment's content and the norms or comparison groups used.

d. Seek to evaluate all appropriate assessment strategies and instruments before selecting a particular instrument.

e. Incorporate standards for educational testing developed by the assessment community and others (including the American Educational Research Association, National Council on Measurement in Education, American Psychological Association, the education agencies of other states, and private education organizations) into the process of evaluating and selecting an assessment instrument.

f. Select an assessment only if potential users have individuals available to them who are, or can be trained to be, capable administrators of the assessment and interpreters of its results.

g. Make information about the strengths and weaknesses of the assessment instruments being evaluated available to interested parties, as long as disclosure does not undermine security precautions outlined elsewhere in this Code.

h. Thoroughly document the process of evaluating and selecting an assessment instrument and make the information available to interested parties prior to its administration.

2. **Select assessments that minimize opportunities for misuse or misinterpretation of results.**

   a. Attempt to identify how interested parties (including state and local policymakers, school officials, parents or the legal representatives of the children assessed, the media, and the general public) will use and interpret the potential results of an assessment.

   b. Consider the positive and negative consequences of such uses and interpretations on the students assessed, the educational units involved, and the community.

   c. Seek information from developers and others on how assessments being considered have been or could be misused, misinterpreted, or overinterpreted.

   d. Select assessment instruments that minimize the potential for misuse and misinterpretation by providing:

      (a) Clear definitions of the appropriate uses and interpretations of the instruments' results.

      (b) Reports that communicate assessment results clearly, at levels of detail appropriate to those receiving them.

      (c) Opportunities for those who will administer and use the assessment results to have input into the selection process.

3. **Insist that assessment developers comply with their responsibilities to those evaluating their instruments for selection.** In order to fulfill their responsibilities developers:

   a. Must accurately represent their instrument and its characteristics, purposes, uses and limitations.

   b. Must ensure that the assessments they produce meet the professional standards of educational assessment.

   c. Must avoid withholding information concerning their assessments, especially where such disclosures would adversely affect the evaluation and selection process.

   d. Must disclose and correct inaccuracies discovered in assessment instruments or their supporting materials as soon as feasible.

e. Should seek evaluation of their instruments from individuals or organizations that are independently recognized as experts in the field of assessment, and make such evaluations available;

f. Should explain relevant concepts and data needed to evaluate their assessment at a level of detail appropriate for those evaluating and selecting assessments.

g. Should disclose previous users of their instrument so that they may be contacted by those evaluating it for use.

h. Should strive to completely and objectively report data on pretesting, standardization, validation, and other steps taken in producing the instrument, including both the positive and negative implications that the data may have for the use of the assessment.

i. Should identify special skills needed to administer or interpret the results of their assessment.

j. Should attempt to minimize the re-use of test formats, items, or tasks.

4. **Recognize and avoid the mixed motives and conflicts of interest that may be present in the development and selection of an assessment:**

   a. Recognize that an instrument's developers have mixed motives that may result in promising more than their assessment instruments can deliver.

   b. Avoid conflicts of interest between those who evaluate and select an assessment and its developer (e.g., an evaluator serves on the board of directors of the developing company or holds substantial financial interest in the company).

   c. Disclose potential conflicts of interest to those responsible for the assessment.

   d. Disclose any attempt by any party to exert undue influence on the evaluation and selection process to those responsible for the assessment.

5. **Establish procedures to ensure that assessments being developed or evaluated are kept confidential in order to minimize disclosure of actual assessment content, questions or their format, or other information which could provide an advantage in student preparation.**

   a. Obtain signed confidentiality agreements from those with access to the assessment and its supporting materials.

b. Limit access to the location where the assessment is being evaluated to only those individuals with a legitimate need.

c. Collect and destroy extra copies of notes or drafts, maintaining only those necessary to continue the development, evaluation, and selection process.

d. Account for all copies of evaluation materials retained once an assessment has been selected.

e. Review all disclosures made pursuant to other parts of this Code in order to minimize the access of those involved in preparing students and administering the assessment to specific information on actual assessment content, questions, and format.

C. **On Preparing Students for an Assessment**

Test preparation is the source of many of the problems that concern us here. It is at this point in the testing process where the pressure to raise test scores comes to bear with the most force, on those individuals who have the least input or power within the process. Teaching to the specific test content is a human, although inappropriate or unethical, response to this situation. The logic is understandable: If educators are to be held accountable for their students learning a particular body of knowledge or set of skills, it is in their best interest to teach those specific things. Indeed, assessment programs are often used by policymakers or administrators as a mechanism to drive the kind of instruction they believe students should receive. The problem is in defining the content to be tested or making sure that the content to be tested is broad enough to prevent narrowing the curriculum taught.

Due to the effects that inappropriate test preparation activities may have on a test's validity—the possibility of test score "pollution," as well as the tendency to narrow the scope of what students actually learn—the question of whether particular practices are appropriate or inappropriate is vital. Unfortunately, the lack of clear lines dividing appropriate and inappropriate practices can lead to many instances where test preparation falls on the inappropriate side of the line. Despite several surveys that document the fact that educators prepare their students using practices that members of the professional assessment community would consider inappropriate or unethical, none of the codes or standards developed so far have directly addressed issues of test preparation. Because any assessment covers only a sample of the content area being assessed, test scores are used to infer mastery of the larger content area being sampled. As instruction is narrowed to what is on the test these inferences become increasingly unjustified.

The provisions in this section were derived from looking for consensus in the current research on preparation for assessments, as well as from the *Code of Fair Testing Practices*, the *Code of Ethical Assessment in Education*, the Coordinator's Manual published by the Michigan Department of Education's Michigan Educational Assessment Program (MEAP

Coordinator's Manual), and Mehrens & Kaminski (1989).

1. **Follow general guidelines for preparing students for an assessment:**

   a. Avoid any preparation that in effect raises assessment scores without simultaneously increasing student mastery of the content domain assessed.

   b. Avoid preparation activities that undermine the accuracy of inferences drawn from the results of the assessment.

   c. Communicate assessment objectives to all involved in preparing students for the test, especially teachers.

   d. Prepare students to master the objectives of the assessment as part of a general overall review, rather than mastering the assessment itself.

   e. Change assessment content periodically in order to focus instruction on the underlying domain, rather than the specific test content.

2. **Avoid test-specific instruction when preparing students for an assessment:**[4]

   a. Do not prepare students through the use of actual questions or tasks found on the assessment or a copy of the current assessment itself.

   b. Do not develop curriculum that is based solely on the content or objectives of an assessment.

   c. Do not limit preparation to the concepts and skills on which students performed poorly in previous assessments.

---

[4] This section warns against instruction designed to raise test scores without raising student mastery of the content. While including the tested content within an overall instructional program is appropriate, rote memorization or practice on specific test content is not.

d. Avoid test preparation activities during the time period[5] immediately preceding an administration of an assessment, other than to teach or review test-taking skills.

e. Avoid limiting preparation for an assessment to questions framed in the format used on the actual assessment.

f. Do not practice or prepare students on published "parallel" forms of the current assessment instrument.

3. **Provide for monitoring and identification of inappropriate preparation activities.**

   a. Establish procedures at each educational unit involved for:

      (1) Educating staff in appropriate methods of preparation for the assessment.

      (2) Monitoring preparation activities, including unannounced observation of classroom preparation activities during periods of time immediately before administration of the assessment.

      (3) Reporting improper preparation activities so that appropriate remedial or, in the most egregious cases, disciplinary action may be taken.

   b. Establish and announce the availability of communication channels allowing teachers, students, and parents or legal representatives to voice their concerns about practices they consider inappropriate.

4. **Follow guidelines on other specific preparation activities:**

   a. Teaching or reviewing test-taking skills is appropriate as long as it does not focus instruction in the content area or format used on the current assessment instrument. For example, multiple- choice, performance assessment, or observation should be employed.

   b. Preparing students for filling out demographic or other preliminary information is appropriate.

---

[5] This period of time is difficult to define here. It should be defined by those administering the assessment on an assessment-by-assessment basis in order to take into account the nature and requirements of the assessment. However, a good rule of thumb is that preparation focused on the assessment itself should decrease as the date of assessment nears.

c. Using commercially prepared score-boosting materials or activities focused specifically on boosting scores, not on improving student knowledge and skills, is inappropriate.

d. Calling attention during review to the fact that a similar question will be on the approaching assessment is inappropriate.

e. Excusing or dismissing students with language or other obstacles to achievement on an assessment is appropriate only when that exclusion is based on current state guidelines or an educationally defensible purpose in the best interest of the individual student.

f. Creating an unnecessary level of apprehension concerning an approaching assessment is inappropriate as the stress may undermine student performance.

## D. On Administering and Scoring Assessments

All who are involved with a testing program have an expectation that it will be implemented with appropriate care. Those who have a stake in the results of an assessment must be sure that they can trust the accuracy of the data that it provides. To that end, all efforts should be taken to see that the administration of an assessment does not undermine its reliability and validity.

The importance of test security needs to be continually emphasized. Breaches in security or deliberate attempts to manipulate test results are serious and should be treated as such. Uniformity and security during the administration of a testing program is a key component in assuring that an assessment program is reliable and useful. If security is lacking many doors may be opened to those whose response to the pressure to perform is to inappropriately raise the test scores of their charges. If the administration of a test is not done uniformly, the inferences and uses for which test developers have validated their assessments may become meaningless as they are often inherently linked to the way in which the test is administered.

The provisions of this section were derived for the most part from the *Code of Fair Testing Practices*, the *Code of Ethical Assessment in Education*, the American Psychological Association's *Standards of Educational and Psychological Testing*, the MEAP Coordinator's Manual, and *Regaining Trust: Enhancing the Credibility of School Testing Programs*.

1. **Be aware of general guidelines on assessment administration:**

   a. Strictly follow administration and scoring procedures prescribed by an assessment's developers.

b. Administer assessments as uniformly as possible among the educational units and specific population(s) assessed.

c. Provide security for assessment materials before, during, and after administration, treating breaches of security as important to the reliability and validity of the assessment's results.

d. If practicable, assessments should be administered or proctored by individuals who have little or no stake in the results.

e. Prior to administering an assessment, provide interested parties (including but not limited to students, their parents or legal representatives, teachers, and school officials) general information on:

(1) The purpose(s) of the assessment

(2) Its general content

(3) The scoring criteria that will be used

(4) How results will be used, reported,and distributed

f. Develop a written policy on the administration of assessments and disseminate it to all individuals who have a role in administering assessments. Such a policy should:

(1) Clearly outline the responsibilities of students, teachers, administrators, and other actors in the administration process.

(2) Identify practices to be followed and avoided in administration of assessments.

(3) Establish security procedures to be followed at each educational unit before, during, and after administration of an assessment.

g. Establish procedures for addressing breaches of the written policy on assessment administration. These procedures should:

(1) Hold accountable all individuals whose actions or pressure to raise test scores formed the basis for the breach.

(2) Rely primarily on education and training as the primary response to a breach, limiting sanctions relating to employment only to the most egregious cases, or to instances of repeated breaches by a particular individual.

(3) Be clearly communicated to all who will be subject to the policy.

h. Develop a written policy on student cheating during an assessment and disseminate it to all who have a role in administering assessments. Such a policy should:

(1) Identify foreseeable or known cheating practices.

(2) Identify conditions in the testing environment that increase the potential for cheating.

(3) Establish monitoring procedures during administration of assessments.

(4) Clearly define action to be taken when cheating is observed or exposed.

i. Prepare an outline of the policy, clearly stating the prohibited practices and actions that will be taken if cheating occurs and communicate it to students and their parents or legal representatives.

2. **Provide testing conditions that will not undermine the reliability and validity of an assessment.**

a. Follow all conditions and instructions prescribed by an assessment's developer as they are closely tied to its validity and reliability, making exceptions only after consultation with the developer, or upon carefully considered professional judgment in the best interests of the children being assessed.

b. Provide an opportunity prior to administration of an assessment for those involved to clarify their understanding of instructions, procedures, and appropriate testing conditions.

c. Administer assessments, especially those that are of the same form, simultaneously throughout all educational units assessed.

d. Make-ups should be handled under similar conditions as soon as possible after the initial administration of the assessment.

e. Ensure that all instructions and test questions are administered uniformly throughout all educational units assessed, unless specifically allowed by the assessment's developer or required by law.

f. Establish and implement procedures for identification of students, seating assignments, space between seats, and other testing conditions.

g. Provide a similar environment for each student assessed, minimizing differences in environment between rooms within a school in which the assessment is given, and between all educational units involved in the assessment as a whole. The assessment

environment should at least be reasonably comfortable, with minimal distractions, avoiding differences in the level of noise or other distractions, extremes in temperature, and amount of working space.

h. All reasonable accommodations should be made, within state guidelines, to ensure that scores of disabled students or students with limited English proficiency are not prejudiced by the way in which an assessment is administered. For example:

(1) Students with visual impairments may need to have instructions and questions read to them, provided in large-print form, or in Braille.

(2) Students with hearing disabilities may need written or signed instructions and questions.

(3) Students whose primary language is not English should not be assessed in English in content areas where their ability to read English is not measured (e.g., assessing mathematic, science or social studies content), unless provided reasonable accommodation.

3. **Limit access to assessment materials before, during, and after its administration**

a. Establish and implement security precautions that minimize access to the assessment materials prior to the time that administration takes place.

b. Adjust access limitations to the character of the particular assessment instrument and the stakes attached to its results; more security is necessary on "high-stakes" assessments than when sampling student achievement.

c. Some specific precautions that can be taken to increase the security of an assessment administration include:

(1) Prior to administration, seal assessment materials in boxes or in shrink-wrap; seal assessment question booklets with gummed labels.

(2) Assessment materials, question booklets, and answer sheets should not be present in schools to be assessed until shortly before administration of the assessment is to begin.

(3) Collect from and return assessment materials to a secure location as soon as administration of the assessment is completed.

(4) Maintain records accounting for the number of assessment booklets and answer sheets distributed and returned.

d. Prepare a written summary of security procedures and communicate it to all involved in assessment administration.

4. **Monitor assessment administration.**

   a. Identify an uninvolved individual at each educational unit assessed to monitor whether administration and security procedures are followed.

   b. Provide for unannounced observation and consultation with those administering an assessment.

   c. Provide for channels through which teachers, students, and parents may communicate concerns about practices they consider to be inappropriate.

5. **Monitor assessment scoring to ensure the reliability of its results.**

   a. Follow all the directions of the developer for scoring the assessment.

   b. Establish procedures for ensuring the accuracy of the scoring process.

   c. Develop auditing procedures to review assessment scoring and overall test results in order to ensure that data are processed according to the established procedures and the developer's instructions.

   d. Develop procedures to maintain security of assessment results until scoring is complete and final reports are issued.

   e. Develop procedures to assist in identifying student cheating and unethical preparation or administration practices. These procedures may include:

      (1) Computer studies of test results to reveal unusual patterns to responses to assessment questions.

      (2) Erasure counts by class or individual school.

      (3) Analysis of patterns of responses from students seated in close proximity.

      (4) Analysis of unusual achievement on the assessment as compared to predicted scores or the previous year's performance.

6. **Be aware of guidelines on other specific administration activities.**

   a. Do not make changes to student answer sheets unless specifically allowed to do so in the developer's administration instructions.

   b. Do not allow anyone but the student tested to transfer responses from a testing booklet to an answer document, unless provided for by the assessment's developer or required by accommodation provided for disabled students.

   c. Assessment administrators should not coach students in any way (e.g., through facial expressions, gestures, or body language) that indicates to the students that their responses may be wrong at any time before or during administration of the assessment.

   d. Assessment administrators should not provide definitions of terms or words used in an assessment unless specifically allowed in developer's instructions.

   e. Assessment administrators should not answer factual questions of students that relate to the content of the assessment questions or tasks.

   f. Except where specifically provided by state law or policy, the scores of a particular educational unit should not be raised by excluding low-scoring students or groups of students from the assessment.

E. **On Interpretation and Use of Assessment Results**

Much of the current criticism of testing programs is associated with the way tests are used and the types of inferences that are drawn from their results. There is widespread agreement among educators, the assessment community, and test publishers that tests are often used for purposes for which they were neither designed nor validated; in addition, their results are often misinterpreted. Despite being selected, prepared for, and administered appropriately, if an assessment's results are interpreted and used inappropriately the validity of the entire exercise has been undermined, or even destroyed.

The provisions of this section were derived for the most part from the *Code of Fair Testing Practices*, the *Code of Ethical Assessment in Education*, and the *Standards of Educational and Psychological Testing*.

1. **Refrain from inappropriate and improper interpretations and uses of assessment results.**

   a. Interpret and use results only after gaining a clear understanding of the assessment administration and scoring, the adequacy of its norms used in its development and other technical features, and its validity for the particular uses contemplated.

   b. Interpret or use the results of an assessment only in a manner specifically recommended and validated by the assessment's developer.

   c. If another use or interpretation of assessment results is required, support must be found to establish the continued validity of the results for that use or interpretation.

   d. Avoid using assessment results to compare or evaluate teachers or administrators of educational units without accounting for other factors that may have influenced differences in results.

   e. Avoid interpretations and uses of the results of an assessment that:

      (1) Compare students or educational units without accounting for the impact of differences in the characteristics of those assessed, as well as disparities in preparation and administration practices.

      (2) Fail to account for differences between the norms or comparison groups used to develop the assessment and the population actually assessed.

      (3) Fail to account for the impact of potential bias, cultural or otherwise, in the content or format of the assessment.

   f. Do not use assessment results as the foundation for claims that cannot be substantiated, or to support false or misleading statements concerning those assessed or the educational units involved.

   g. Avoid using assessment results to justify decisions made primarily on other grounds, such as political pressures, funding considerations, or other noneducational factors.

   h. Ensure that those who are ultimately responsible for an assessment are advised of potential misuses and misinterpretations of its results, so that they may take appropriate action.

2. **Promote appropriate interpretation and use of an assessment's results by encouraging valid inferences.**

   a. The inferences that are drawn from an assessment must be valid for the assessment's

intended uses.

b. Consider assessment results within the context of the educational environment surrounding each educational unit assessed.

c. Avoid interpreting assessment results as a fixed and unchangeable index of student performance.

d. Do not base decisions that will have important effects on individual students, educators, or institutions solely on the results of a single assessment.

e. An assessment may be used appropriately as the end point in a decision-making process, or "gateway," as long as it comes after a sequence of consideration of and decisions about other relevant criteria or indicators of student performance.

f. Intended uses of an assessment should be reevaluated and validated whenever substantial changes are made in format, content, instructions, language, or administration.

3. **Provide information to students and their representatives about access to the assessment results and steps taken to protect their privacy.**

   a. Advise students and their representatives where assessment results will be kept on file and how to gain access to them.

   b. Communicate to students and their representatives information on their individual rights concerning access to the assessment results and questioning the accuracy of obtained scores, as well as how those rights can be exercised.

   c. Establish procedures to protect the privacy of the students assessed and communicate those protections to the students and their parents or legal representatives.

4. **Ensure that test developers fulfill their obligation to assist in proper communication of assessment results.** They should provide:

   a. Simple score reports that describe assessment performance clearly and accurately, especially those that will be provided to parents or legal representatives of the students assessed.

   b. Explanations of the meaning and limitations of assessment scores, the populations representing any norms used, the process used to select samples of those assessed, and the age of these data.

   c. Information on foreseeable misuses of an assessment's results.

d. A reasonable and appropriate means for setting passing scores.

5. **Identify all audiences to which the results of an assessment may be communicated and educate them in appropriate interpretations and uses of its results.**

   a. Ensure that explanations of an assessment are appropriate for the level of understanding of each audience identified.

   b. Provide each audience with background information about the assessment, including:

      (1) Its intended purposes and uses.

      (2) How the results were derived, and how scores and scoring summaries were developed and their proper interpretation.

   c. Acknowledge and be prepared to justify the potential direct and indirect consequences that assessment results may have on individuals, educational units, and educational programs.

6. **Minimize potential misinterpretation and misuse when reporting assessment results to students, parents, legal representatives, teachers, the general public, and the media.**

   a. Provide all audiences requesting assessment results with clear descriptions of the assessment measures, what scores mean, common misinterpretations, and how the scores will be used, at an appropriate level of understanding.

   b. Challenge misinterpretations, invalid comparisons, and other misuses of assessment results by the media or other parties with an effort to educate them concerning the limits of the inferences that may be drawn from assessment results.

   c. Respond to misinterpretations of assessment results in the media through the same media in which the misinterpretation appeared.

7. **Communicate the results of an assessment, and interpretations of the results, within the context of the assessment's limitations.** Be sure to include:

   a. The shortcomings of the type or quality of the particular instrument used in assessing the content involved.

   b. The impact that characteristics of the students assessed has on the assessment's reliability and/or validity.

   c. The adequacy of the norms or standards used in interpreting the assessment's results.

*Conclusion*

Blame often seems easy to dispense when inappropriate or unethical practices are employed to raise scores on assessments, or when their results are misused. There are always those at whom a finger can be pointed, whether they are teachers who believe that they are "saving their job," administrators "promoting" their schools or districts, a politician wanting attention, or others acting in self-interest. However, the roots of inappropriate or unethical testing practices are much more complex. The roots may be fed by any of the individuals involved with assessments, including those who develop the tests themselves, the policymakers and administrators who choose assessments and interpret or act upon their results, or the educators who prepare students for tests and administer them. The roles of each of these actors, and the means and motivations by which they play their parts, all may have an impact on whether or not assessments are appropriately and effectively conducted, and whether their results can be relied upon.

Where testing programs are selected, administered, and used appropriately, there is no doubt that they can maintain a valuable position in American education. However, until consensus allows American education to move away from what some see as its excessive reliance on large-scale assessments to provide the information required to make informed educational decisions, such testing programs will continue to hold their dominant role. The reality of inappropriate or unethical assessment practices, and the real damage they cause to the educational process, must be confronted. This Code is intended to help educators and policymakers to confront this and guide them in understanding and making appropriate choices in large-scale assessments, thus ensuring an accurate and fair foundation for decisions based upon them.

## APPENDIX A

### DEFINITION OF TERMS

**Assessment:** Student assessment refers to any method of determining what students know and can do, including testing, teacher observations, collections of students' work (portfolios), projects, and performances. Distinctions are made between traditional assessments, which require students to select a right answer, and performance assessments, which require students to create an appropriate response to a performance task. The most common traditional assessments include multiple-choice, true-false, and matching tests. Performance assessments can range from essay exams to situations where students must build a model or design and perform an experiment. In the first case, student performance is inferred from the test score, while in the second case, the performance is directly observed.

**Assessment materials:** Materials gathered during the development, evaluation, and selection process, as well as the chosen assessment's question and answer booklets, and other materials that disclose information that might compromise the validity and reliability of the assessment.

**Bias:** Any influence on test scores that is due to the prejudicial treatment of certain groups of test takers.

**Comparison groups:** In a norm-referenced assessment, a student's scores are compared against those of a national group of students at the same grade level.

**Conflict of interest:** Usually involves a conflict between the "public" responsibility of those involved in the development and selection process and their private pecuniary interests.

**Content:** The subject matter, knowledge, and skills being assessed by an assessment instrument.

**Content domain:** The totality of the content that could be assessed.

**Content sample:** The portion of the content domain that is assessed in a particular assessment instrument.

**Developers:** Commercial vendors and state or local assessment producers, either private consultants or state employees.

**Educational units:** The classroom, school, district, and state level.

**High Stakes Assessment:** Any assessment where the results lead to rewards or sanctions for individual educators or institutions.

**Inferences:** Interpretations made about a student or institution based upon performance on an assessment.

**Instruments:** The actual assessment, its question and answers, and supporting documentation.

**Large-Scale Assessments:** Standardized tests, including commercial norm-referenced tests, tests developed at the district or state level, state-developed assessment programs, and federally imposed assessments such as the NAEP. Because results are used across many different settings, the comparability of scores across settings is very important.

**Norms:** The reference scores of a representative sample of students against which a test-taker's scores are compared.

**Objectives:** A set of skills and content knowledge included on an assessment.

**Reliability:** A measure of the consistency of an individual score regardless of when, where, and by whom the student is assessed.

**User:** The officials of educational units who will interpret and use the results to inform their decisionmaking.

**Validity:** An indication of how well an assessment actually measures what it purports to measure and provides accurate information to support inferences made.

## APPENDIX B

## BIBLIOGRAPHY

In preparing this model policy code the following sources provided a wealth of information on the ethical issues involved in the practice of large-scale assessment, as well as specific appropriate and inappropriate assessment practices.


Bell, G. V. (1993). *The test of testing: Making appropriate and ethical choices in assessment.* Oak Brook, IL: North Central Regional Educational Laboratory.

Cannell, J. J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average.* Daniels, WV: Friends for Education.

Cannell, J. J. (1989). *The "Lake Wobegon" report: How public educators cheat on standardized achievement tests.* Albuquerque, NM: Friends for Education.

Cannell, J. J. (1990). *Testing ethics model legislation.* Paper presented at annual meeting of the National Council on Measurement In Education, Boston, MA.

*Code of ethical assessment practices in education.* (Draft #1, February 1993). National Council On Measurement In Education Ad Hoc Committee on the Development of a Code of Ethics.

*Code of fair testing practices in education.* (1988). Washington, DC: Joint Committee on Testing Practices.

Cuban, L. (1991). *The misuse of tests in education.* Paper prepared for the Office of Technology Assessment, September, 1991.

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution. *Educational Researcher, 20*(5), 2-7.

Hall, J. L., & Kleine, P. F. (1990). *Educators' perceptions of achievement test use and abuse: A national survey.* Paper presented at the annual meeting of the National Council on Measurement In Education, Boston, MA.

Koretz, D. (1988). Arriving in Lake Woebegone: Are Standardized Achievement Tests Exaggerating Advancement and Distorting Instruction? *American Educator, 12*, 2.

Linn, R. L. (1991). *Test misuse: Why is it so prevalent?* Boulder, CO: Center for Research on Evaluation, Standards, and Student Testing.

Linn, R., Graue, M., & Sanders, N. (1989). *Comparing state and district test results to national norm: Interpretation of scoring "Above the National Average."* Paper presented at the annual meeting of the American Educational Research Association, March, 1989, Chicago, IL.

Madaus, G. F. (1985). Test scores as administrative mechanisms in educational policy. *Phi Delta Kappan, 66*, 611-617.

Madaus, G. F. (1990). The distortion of teaching and testing: High-stakes and instruction. *Peabody Journal of Education, 65*, 29-46.

Mehrens, W. A. (1991). *Defensible/indefensible instructional preparation for high stakes achievement tests: An exploratory trialogue.* Revision of presentation given in a symposium of the same title at 1991 AERA/NCME Annual Meeting, April 10, 1991, Chicago, IL.

Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice, 8*, 14-22.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement, 21*, 215-37.

NCME Task Force (1991). *Regaining trust: Enhancing the credibility of school testing programs.* A report by a task force of the National Council on Measurement in Education, April 1991.

Nolen, S. B., Haladyna, R. M., & Haas, N. S. (1990). *A survey of actual and perceived uses, test preparation activities, and effects of standardized achievement tests.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

*Testing in American schools: Asking the right questions.* Office of Technology Assessment Report.

Phillips, G. W., & Finn, C.E. (1988). The Lake Woebegone effect: A skeleton in the testing closet? *Educational Measurement: Issues and Practice, 7*, 2.

Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice, 10*(4), 12-15.

Shepard, L. (1989). *Inflated test score gains: Is it old norms or teaching to the test?* Paper presented at the annual meeting of the American Educational Research Association, March, 1989, San Francisco, CA.

Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal, 3*, 521-42.

Suarez, T. M., & Gottovi, N. C. (1992). The impact of high stakes assessment on our schools. *NASSP Bulletin, 76*, 82-88.

**North Central Regional Educational Laboratory**
1900 Spring Road, Suite 300
Oak Brook, IL 60521-1480
(708) 571-4700
Fax (708) 571-4716

BEST COPY AVAILABLE